UNIVERSIDAD NACIONAL AGRARIA LA MOLINA

DEPARTAMENTO DE ESTADÍSTICA E INFORMÁTICA
FACULTAD DE ECONOMÍA Y PLANIFICACIÓN

TRABAJO INTEGRADOR

Inferencia Estadística 2025-II

Integrante	Código
Montúfar Paiva Yeraldi Mercedes	20230400
Kay Daniela L. Zavala Malpartida	20230420
Rojas Taco Fabiana	2023
Castillo Ruiz Mauricio Gabriel	20230384
Gómez Vigo Héctor Estefano	20230397

Docente: Fernando Miranda Villagómez

LA MOLINA - LIMA - PERÚ 2025

Distribución Pareto:

$$f(x) = \frac{\beta\alpha^\beta}{x^{\beta+1}} I_{(\alpha,\infty)}(x)$$

$$E(X) = \frac{\alpha\beta}{\beta-1} \quad ; \quad \text{VAR}(X) = \left(\frac{\alpha}{\beta-1}\right)^2 \frac{\beta}{\beta-2}$$

1.1 EMV para parametro alfa:

Usaremos el Método de máxima verosimilitud para poder hallar el primer estimador para alfa:

$$L(\alpha,\beta) = \prod_{i=1}^{n} f(x_i;\alpha,\beta) = \prod_{i=1}^{n} \frac{\beta\alpha^\beta}{x_i^{\beta+1}} I_{[\alpha,\infty)}(x_i) \quad = \beta^n \alpha^{n\beta} \left(\prod_{i=1}^{n} x_i^{-(\beta+1)}\right) \times \prod_{i=1}^{n} I_{[\alpha,\infty)}(x_i)$$

Analizando la indicadora:

$$\begin{aligned}\prod_{i=1}^{n} I_{[\alpha,\infty]}(x_i) &= I\left(\bigcap_{i=i}^{n} \{x_i \geq \alpha\}\right) \\ &= I(\alpha \leq x_1, \alpha \leq x_2, \ldots \alpha \leq x_n) \\ &= I(\alpha \leq mi(x_1,\ldots,x_n\}) = I(\alpha \leq y_1]\end{aligned}$$

donde y1 = X(1) es el mínimo de la muestra:

$$El \text{ EMV de } \theta = \alpha \text{ es } \ T = Y_1.$$

1.1.2 Propiedades:

a) Insesgamiento: Distribución del mínimo:

$$X_1 \ldots X_n \sim \text{ Pareto } (\alpha,\beta) \text{ i i d}$$

$$F(x) = 1 - \left(\frac{\alpha}{x}\right)^\beta, x \geqslant \alpha$$

$$y = x_i$$

$$Fx_i(y) = n[1-F(y)]^{n-1}\mathcal{F}(y) \to n\left[\left(\frac{\alpha}{\beta}\right)^\beta\right]^{\beta-1} \cdot \frac{\beta\alpha^\beta}{y^{\beta+1}}$$

$$= n\left[\frac{\alpha^{\beta(n-1)}}{\alpha^{\beta(n-1)}}\right] \cdot \frac{\beta\alpha^\beta}{y^{\beta+1}} = \frac{(n\beta)\alpha^{n\beta}}{y^{n\beta+1}}, y \geqslant \alpha$$

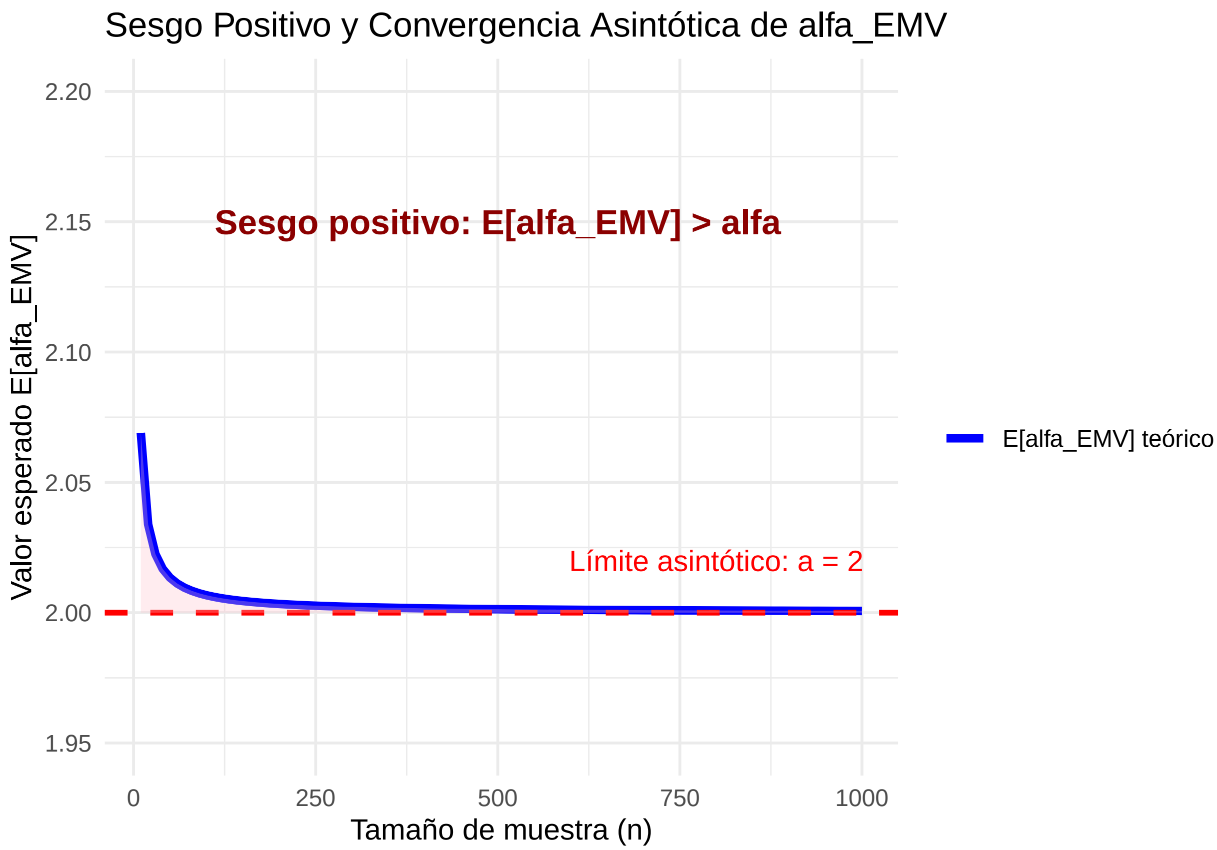
$$\to x_1 \sim \text{Pareto}(\alpha, n\beta)$$

Hay sesgo positivo, es decir que sobrestima al parametro, no cumple propiedades de Insesgamiento. Ahora lo veremos mejor con los valores para nuestra data:

```
# 1. Valor observado del mínimo
alpha_EMV_observado <- min(reclamos)
# 2. Valor esperado teórico del mínimo
E_alpha_EMV_teorico <- (n * beta_true * alpha_true) / (n * beta_true - 1)
```

```
## [1] "El valor observado es: 2.00031032898679 y el valor esperado es: 2.00066688896299"
```

El valor observado en los datos fue de 2.00031, mientras que el valor esperado teórico es 2.00067, ambos ligeramente por encima del valor real del parámetro (alfa = 2.00000). Esto confirma que, para muestras finitas, el estimador tiende a sobrestimar el parámetro verdadero, aunque en este caso la magnitud del sesgo es muy pequeña (del orden de 0.0003 a 0.0007). La cercanía entre el valor observado y el teórico valida la expresión matemática del sesgo y respalda la propiedad de insesgamiento asintótico, ya que a medida que el tamaño de muestra aumenta, el sesgo tiende a cero.



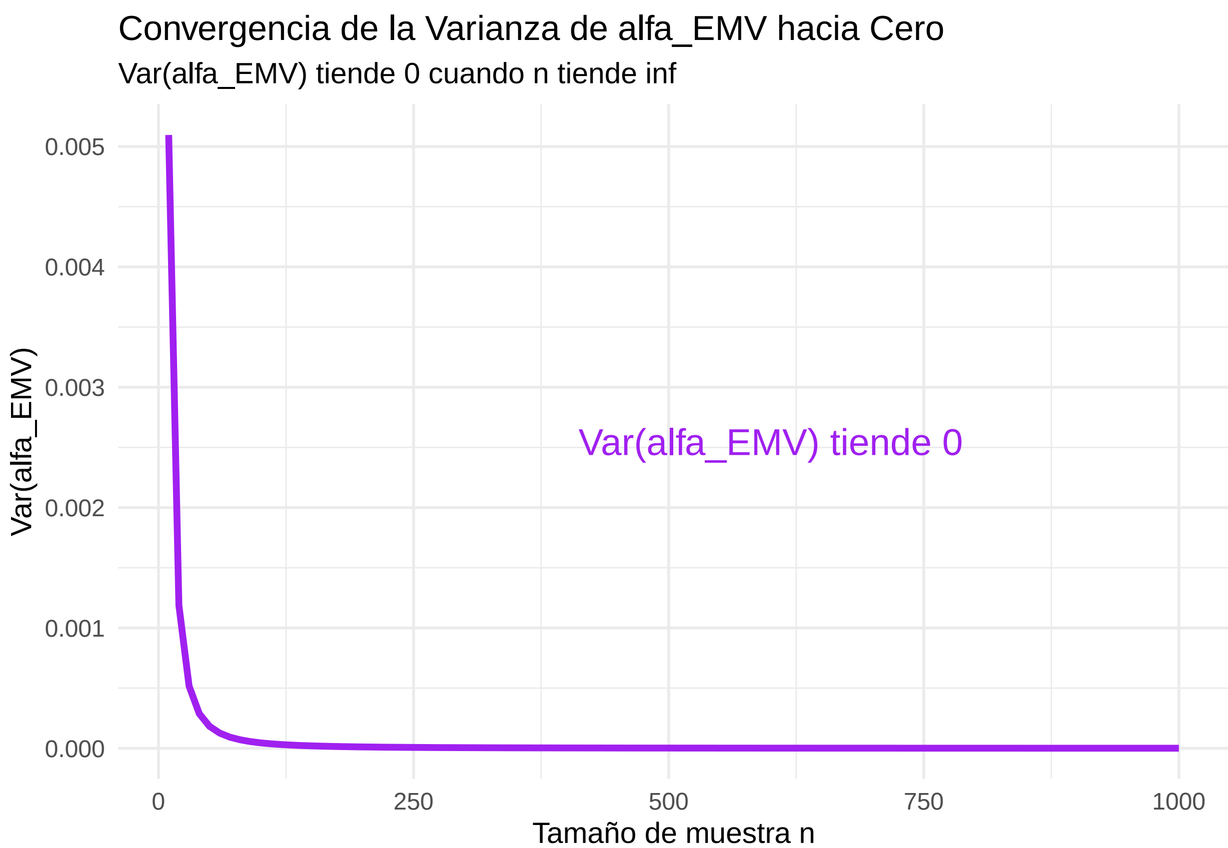
de nuestra base de datos, tomamos submuestras que van incrementando su valor para verificar que tendiendo al infinito (tamaños suficientemente grande) podemos ver como el valor tiende al parámetro de alfa = 2.

b) Consistencia: Podemos hacer uso del Teorema 2, por la propiedad anterior del insesgamiento asintotico

- $\lim_{n\to\infty} E(y_1) = \alpha$ ya que $\lim_{n\to\infty} \frac{n\beta \cdot \alpha}{n\beta - 1} = \lim_{n\to\infty} \frac{\beta\alpha}{\beta - \frac{1}{n}} = \frac{\beta\alpha}{\beta} = \alpha$
- $\lim_{n\to\infty} \mathrm{Var}(y_1) = \lim_{n\to\infty} \frac{n\beta\alpha^2}{(n\beta-1)^2(n\beta-2)} = \lim_{n\to\infty} \frac{n\beta\alpha^2}{n^3\beta^3} = \lim_{n\to\infty} \frac{\alpha^2}{n^2\beta^2} = 0$

Como podemos obervar cumple ambas condiciones, por ende decimos que nuestro estimador es consitente n otras palabras el estimador converge en probabilidades al parametro.

** Gráfico:** Hemos visto ya un gráfico para solidar la primera demostración, ahora veamos uno que consolide la segunda condición:



c) Suficiencia:

$$P(x = x/t = T) = \frac{P\left(x_1 = x_1 \ldots x_n = x_n, T = y_1\right)}{P(T = y_1)}$$

$$\frac{\left(\frac{\beta\alpha^\beta}{x_1^{\beta+1}}\right)\left(\frac{\beta\alpha^\beta}{x_2^{\beta+1}}\right)\cdots\left(\frac{\beta\alpha^\beta}{x_n^{\beta+1}}\right)}{\frac{n\beta\alpha^{n\beta}}{y^{n\beta+1}}}, \quad x_i \geqslant \alpha,\ i = 1, 2, \ldots, n$$

$$= \frac{\beta^n \cdot \alpha^{n\beta} \cdot \prod_{i=1}^n x_i^{-(\beta+1)}}{\frac{n\beta\alpha^{n\beta}}{y^{n\beta+1}}} = \frac{\beta^{n-1} \cdot y^{n\beta+1} \cdot \prod_{i=1}^n x_i^{-(\beta+1)}}{n}$$

Dada la dsitribución condicional dado Y = X1, no depende de alfa, entonces podemos decir que es una estadística suficiente para alpha, cumple la propiedad.

d) Ancilaridad: La distribución del mínimo $x_{(1)}$ es:

$$x_{(1)} \sim \text{Pareto}(\alpha, n\beta)$$

$$f_{x_{(1)}}(t) = \frac{n\beta\alpha^{n\beta}}{t^{n\beta+1}}, \quad t \geqslant \alpha$$

Como vemos su distribución depende de alfa, no cumple con la propiedad de ancilaridad.

e) Completitud: Calculamos la esperanza de $g(T)$:

$$E[g(T)] = \int_{\alpha}^{\infty} g(t) \cdot \frac{n\beta\alpha^{n\beta}}{t^{n\beta+1}} \, dt.$$

Exigimos que $E[g(T)] = 0$ para todo $\alpha > 0$:

$$\int_{\alpha}^{\infty} \frac{g(t)}{t^{n\beta+1}} \, dt = 0 \quad \forall \alpha > 0.$$

Derivamos ambos lados respecto a α:

$$\frac{d}{d\alpha}\left[\int_{\alpha}^{\infty} \frac{g(t)}{t^{n\beta+1}} \, dt\right] = -\frac{g(\alpha)}{\alpha^{n\beta+1}} = 0.$$

Esto implica que:

$$g(\alpha) = 0 \quad \forall \alpha > 0.$$

Como α es cualquier valor positivo, concluimos que $g(t) = 0$ para todo t en el soporte de T. Por lo tanto, $T = X_{(1)}$ es una **estadística completa** para α cuando β es conocido.

Por el Teorema de Bahadur, como es suficiente y completo, podemos afirmar que nuestro estimador es minimal suficiente.

f) Optimalidad: En nuestro caso alfa_EMV no es óptimo porque es sesgado. Pero si existe un estimador óptimo basado en la estadística suficiente que se podría obtener mediante corrección del sesgo. (Teorema de Lehman-Scheffé)

1.2 EMV para parametro Beta:

Para hallar el estimador debemos remplazar nuestro primer estimador para el parametro alfa:

$$\hat{\alpha}_{MLE} = \min\{X_1, X_2, \ldots, X_n\}$$

Como sabemos el máximo en L(alpha,beta) es igual al máximo de Ln(Lalpha,beta)) entonces haremos uso del ln:

$$L(\alpha, \beta) = \beta^n \cdot \alpha^{n\beta} \cdot \prod_{i=1}^{n} X_i^{-(\beta+1)}$$
$$\ln(L(\alpha, \beta)) = n \cdot \ln(\beta) + n\beta \ln(\alpha) - (\beta + 1) \sum_{i=1}^{n} \ln(X_i)$$

$$
\begin{aligned}
\frac{d \ln(L(\alpha,\beta))}{d\beta} &= \frac{n}{\beta} + n\ln\alpha - \sum_{i=1}^{n} \ln X_i \\
&= \frac{n}{\beta} + n\ln\alpha - \sum_{i=1}^{n} \ln X_i = 0 \\
&\Rightarrow \frac{n}{\beta} = \sum_{i=1}^{n} \ln X_i - n\ln\alpha \\
&\Rightarrow \frac{n}{\beta} = \sum_{i=1}^{n} \ln\left(\frac{X_i}{\alpha}\right) \\
&\Rightarrow \hat{\beta} = \frac{n}{\sum_{i=1}^{n} \ln\left(\frac{X_i}{x_1}\right)}
\end{aligned}
$$

1.1. Propiedades:

a) Insesgamiento Primero analizaremos una parte del denominador para determinar su distribución y se facilite el procedimiento.

$$
\begin{aligned}
F_Y(y) = P(Y_i \le y) &= P\left(\ln\left(\frac{X_i}{\alpha}\right) \le y\right) = P(X_i \le \alpha e^y) \\
&= 1 - \left(\frac{\alpha}{\alpha e^y}\right)^{\beta} = 1 - e^{-\beta y}, \quad y \ge 0 \\
f_Y(y) &= \frac{d}{dy} F_Y(y) = \beta e^{-\beta y}, \quad y \ge 0
\end{aligned}
$$

Analizando el resultado de la distribución, notamos que tiene la forma de la exponencial. Entonces el mínimo igual será exponencial con parametro n por beta. Ademas tomamos a T como la sumatoria, donde por propiedad de ln de una división podemos desplegarlo.

$$
\begin{aligned}
Y_i &\sim \text{Exponencial}(\beta) \\
W = Y_{(1)} &= \min(Y_1, \ldots, Y_n) \sim \text{Exp}(n\beta) \\
T &= \sum_{i=1}^{n} (Y_i - Y_{(1)}) \sim \text{Gamma}(n-1, \beta) \\
\hat{\beta} &= \frac{n}{T}, \quad T \sim \text{Gamma}(n-1,\beta) \\
E\left[\frac{1}{T}\right] &= \frac{\beta}{n-2}, \quad n > 2 \\
E[\hat{\beta}] &= n \cdot E\left[\frac{1}{T}\right] = \frac{n\beta}{n-2}
\end{aligned}
$$

Acá vemos que hay un sesgo presente

$$
\text{Sesgo} = E[\hat{\beta}] - \beta = \frac{n\beta}{n-2} - \beta = \frac{2\beta}{n-2}
$$

Por ende, no es un estimador insesgado. Ademas al ser positivo podemos decir que sobrestima al parametro.

1.3 Estimador para la Media Poblacional (μ):

Consideramos la media muestral como estimador natural del valor esperado de la distribución:

$$\bar{X} = T(X_1, \dots, X_n) = \frac{1}{n}\sum_{i=1}^{n} X_i$$

Para que los momentos existan en una distribución Pareto, debemos asumir que $\beta > 1$ (para la esperanza) y $\beta > 2$ (para la varianza). El parámetro a estimar es:

$$\mu = E(X) = \frac{\beta\alpha}{\beta - 1}$$

1.3.1 Propiedades:

a) Insesgamiento: Calculamos el valor esperado del estimador:

$$\begin{aligned} E[\bar{X}] &= E\left[\frac{1}{n}\sum_{i=1}^{n} X_i\right] \\ &= \frac{1}{n}\sum_{i=1}^{n} E[X_i] \\ &= \frac{1}{n} \cdot n \cdot \left(\frac{\beta\alpha}{\beta - 1}\right) = \frac{\beta\alpha}{\beta - 1} = \mu \end{aligned}$$

El estimador es **estrictamente insesgado** para μ. Verificamos esto con la base de datos:

```r
# 1. Valor observado de la media muestral
mu_muestral_obs <- mean(reclamos)

# 2. Valor esperado teórico (mu)
mu_teorico <- (beta_true * alpha_true) / (beta_true - 1)
```

```
## [1] "La media muestral observada es: 2.96908 y el valor esperado teórico es: 3"
```

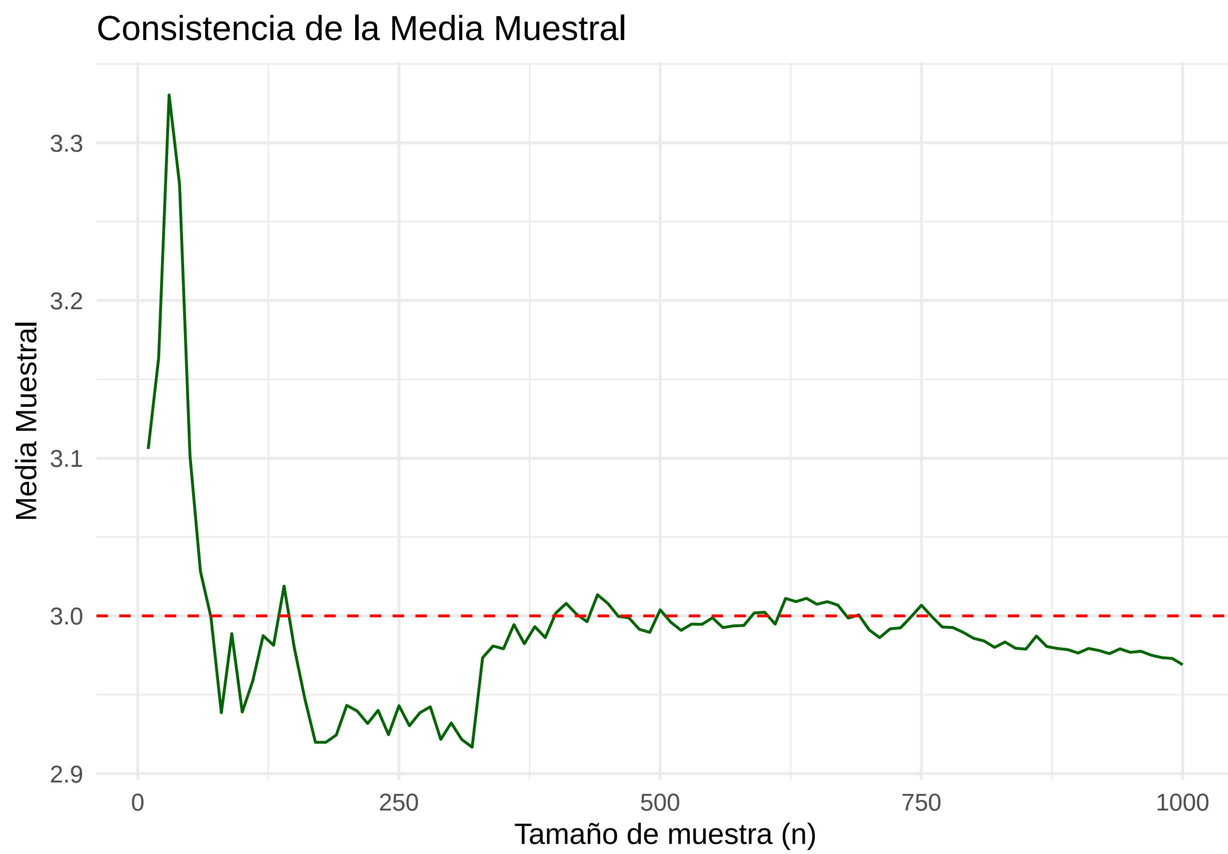
b) Consistencia: Para demostrar la consistencia, verificamos las condiciones del Teorema de Chebyshev (asumiendo $\beta > 2$):

1. **Insesgadez:** $\lim_{n\to\infty} E(\bar{X}) = \mu$ (ya demostrado).
2. **Varianza tiende a cero:**

$$\begin{aligned} \text{Var}(\bar{X}) &= \frac{\text{Var}(X)}{n} \\ &= \frac{1}{n}\left[\left(\frac{\alpha}{\beta - 1}\right)^2 \frac{\beta}{\beta - 2}\right] \end{aligned}$$

$$\lim_{n\to\infty} \text{Var}(\bar{X}) = \lim_{n\to\infty} \frac{C}{n} = 0$$

Al cumplirse ambas condiciones, $\bar{X}$ es un estimador **consistente** para μ.



c) Suficiencia: Aplicamos el Teorema de Factorización de Fisher-Neyman a la función de verosimilitud:

$$L(\alpha, \beta) = \underbrace{\beta^n \alpha^{n\beta} \cdot I_{(\alpha,\infty)}(x_{(1)})}_{g(x_{(1)},\alpha,\beta)} \cdot \underbrace{\left(\prod_{i=1}^{n} x_i\right)^{-(\beta+1)}}_{h(x_1,\ldots,x_n)}$$

Como se observa, los estadísticos suficientes conjuntos para (α, β) son el mínimo $X_{(1)}$ y el producto $\prod X_i$ (o equivalentemente $\sum \ln X_i$). La media muestral $\bar{X} = \frac{1}{n}\sum X_i$ **no puede factorizarse** de manera que contenga toda la información de los parámetros.

Por lo tanto, $\bar{X}$ **no es un estadístico suficiente** para los parámetros de la distribución Pareto.

d) Ancilaridad: Un estadístico es ancilar si su distribución no depende de los parámetros. La distribución de $\bar{X}$ para una Pareto no tiene una forma cerrada sencilla (es una suma de variables Pareto), pero su esperanza $E(\bar{X}) = \frac{\alpha\beta}{\beta-1}$ y su varianza dependen directamente de α y β.

Al depender sus momentos (y por ende su distribución) de los parámetros, el estimador **no es ancilar**.

e) Completitud: Dado que $\bar{X}$ no es un estadístico suficiente para la familia Pareto, no se suele analizar su completitud como estimador. Sin embargo, sabemos que el estadístico suficiente conjunto $(X_{(1)}, \sum \ln X_i)$ es completo, pero la suma aritmética $\sum X_i$ no lo es bajo esta estructura de familia no exponencial (en el sentido de los parámetros naturales de la Pareto).

f) Optimalidad: Un estimador es óptimo (UMVUE) si es insesgado y su varianza alcanza la Cota Inferior de Cramér-Rao (CICR) o si es función de un estadístico suficiente y completo.

1. **Eficiencia:** Al no ser función del estadístico suficiente $(X_{(1)}, \sum \ln X_i)$, la media muestral pierde información.
2. **Comparación:** Existe otro estimador para μ, basado en los EMV de α y β ($\hat{\mu} = \frac{\hat{\beta}\hat{\alpha}}{\hat{\beta}-1}$), que asintóticamente tiene menor varianza que $\bar{X}$.

Conclusión: $\bar{X}$ **no es un estimador óptimo** para la media de una población Pareto, aunque sea fácil de calcular e insesgado.